Hornworts reveal a spatial model for pyrenoid-based CO_2-concentrating mechanisms in land plants

Tanner A. Robison[1,2], Zhen Guo Oh[2], Declan Lafferty[1], Xia Xu[1], Juan Carlos A. Villarreal[3], Laura H. Gunn[2]*, Fay-Wei Li[1,2]*

[1] Boyce Thompson Institute, Ithaca, NY 14853, USA
[2] Plant Biology Section, Cornell University, Ithaca, NY 14853, USA
[3] Department of Biology, Laval University, Quebec City, QC, Canada
*Corresponding authors

ABSTRACT

Pyrenoid-based CO_2-concentrating mechanisms (pCCMs) turbocharge photosynthesis by saturating CO_2 around Rubisco. Hornworts are the only land plants with a pCCM. Owing to their closer relationship to crops, hornworts could offer greater translational potential compared to the green alga Chlamydomonas, the traditional model for studying pCCM. Here we report the first thorough investigation of a hornwort pCCM using the emerging model *Anthoceros agrestis*. The pyrenoids in *A. agrestis* exhibit liquid-like properties similar to Chlamydomonas, but differ by lacking starch sheaths and being enclosed by multiple thylakoids. We found that the core pCCM components in Chlamydomonas, including BST, LCIB, and CAH3, are conserved in *A. agrestis* and likely have similar functions based on their subcellular localizations. Therefore, the underlying chassis for concentrating CO_2 might be shared between hornworts and Chlamydomonas, and ancestral to land plants. Our study presents the first spatial model for pCCM in a land plant, paving the way for future biochemical and genetic investigations.

MAIN

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the carbon-fixing enzyme, is the gatekeeper for virtually all biologically available carbon. Despite its central importance in global primary productivity, Rubisco is considered to have two major limitations: a low rate of activity and poor specificity for carbon dioxide (CO_2) [1]. Photosynthetic organisms have largely overcome the first of these limitations by simply making more of the enzyme, so much so that Rubisco is considered to be the most abundant enzyme in the biosphere [2]. Rubisco's poor specificity means that it can also react with oxygen (O_2), resulting in photorespiration which costs energy and leads to the net loss of fixed CO_2 [3,4].

To reduce photorespiration, some plants have evolved systems to concentrate inorganic carbon (Ci) around Rubisco (called CO_2-concentrating mechanisms; CCM), either through biophysical or biochemical means [5]. Biochemical CCMs (C_4 or CAM photosynthesis) use enzymatic pathways to concentrate carbon, either in vacuoles or in bundle sheath cells. Biophysical CCMs, seen in cyanobacteria, algae, and hornworts, concentrate Ci in subcellular compartments (pyrenoids or carboxysomes) where Rubisco is highly abundant [6].

The model organism for studying pyrenoid-based CCM (pCCM) is the green alga *Chlamydomonas reinhardtii*. Chlamydomonas pyrenoids are a liquid-like proteinaceous compartment whose phase separation is mediated by the interaction of Rubisco and the linker protein EPYC1 [7–9]. In the algal pCCM, a series of Ci channels, such as LCI1, LCIA, and bestrophin channels (BST), facilitate Ci transport from the extracellular space into specialized thylakoid tubules [10]. These tubules traverse pyrenoids and contain a specific carbonic anhydrase (CA), CAH3. This CA catalyzes the conversion of bicarbonate (HCO_3^-) into CO_2, which can freely diffuse out of the thylakoid tubules and into the pyrenoid. A starch sheath surrounds pyrenoids and might function as a CO_2 diffusion barrier to enhance the efficiency of the CCM [11,12]. Another CA, LCIB, localizes around the gaps of the starch sheath to recapture leaked CO_2 by converting it back to HCO_3^-. Recent modeling work suggests that a minimally functional pCCM requires the joint operation of BST, LCIB, and CAH3 (or the equivalent thereof), in addition to a pyrenoid enclosed by diffusion barriers [13].

Installing a pCCM into crop plants may boost CO_2 fixation by as much as 60% [4,14,15]. However, transplanting an algal CCM into land plants is complicated by the fact that around one billion years of evolution separate these lineages [16], over which time significant differences have accumulated in chloroplast structure and protein sequences. On the other hand, hornworts, a group of bryophytes (Fig. 1A,B), are the only known land plants with a pyrenoid or biophysical CCM of any kind [17–19]. Hornwort pyrenoids are functionally analogous to algal pyrenoids, acting as a locus of Rubisco accumulation, and the focal point of the pCCM [17,18,20]. Characterizing the land plant pCCM may provide significant translational advantages. However, virtually nothing is known about the functional components enabling a pCCM in hornworts.

In this study, we characterized the morphology and physical properties of hornwort pyrenoids, as well as proteins involved in the pCCM, using the emerging model hornwort *Anthoceros agrestis* [21–23]. Diffusion properties of fluorescently tagged proteins and live-cell imaging indicate that *A. agrestis* pyrenoids are liquid-like, similar to the Chlamydomonas pyrenoid. Putative CCM components for protein subcellular localization studies were selected by leveraging both Rubisco co-immunoprecipitation (co-IP) and 41 genomes spanning plant and algal diversity. We found that hornworts possess orthologs of several core pCCM components

from Chlamydomonas, and provide evidence for shared functional roles. We thus infer that the chassis for pCCM was present in the last common ancestor of land plants. Comparative genomics and co-IP did not reveal any clear EPYC1 homolog or analog, implying that hornworts might have adopted a different strategy for pyrenoid formation. Based on our findings, we propose the first spatial model for a land plant pCCM, which is consistent with reaction-diffusion modeling for a functional pCCM, and set the stage for future biochemical and genetic investigation.

RESULTS

A. agrestis contains multiple pyrenoids that are not enclosed by starch sheaths

We first characterized the morphology of *A. agrestis* pyrenoids using transmission electron microscopy (TEM). Compared to other hornwort species [17,20], *A. agrestis* pyrenoids are slightly larger in size, with a diameter ranging from 500 nm to 5 µm (Fig. 1c-e; Supplementary Fig. 1). *A. agrestis* (and hornworts in general) have multiple pyrenoids per chloroplast, though the number of which appears to be highly variable (Fig. 1c,d; Supplementary Fig. 1). This pattern differs from the singular pyrenoid found in most algal species including Chlamydomonas [24]. Further, we found a clear lack of any kind of starch sheath surrounding the pyrenoids of *A. agrestis* (Fig. 1d,e), which in Chlamydomonas is believed to act as a critical diffusion barrier to prevent CO_2 leakage [11,12]. While no starch sheath is present, *A. agrestis* has many layers of stacked thylakoids wrapping around the pyrenoids (Fig. 1e; white arrowhead). Such stacked thylakoids might be sufficient to prevent CO_2 leakage, as suggested by recent modeling of the Chlamydomonas pCCM [13].

A. agrestis pyrenoids are liquid-like

While we have previously demonstrated that the pyrenoids of *A. agrestis* are the sites of Rubisco accumulation [23], the physical properties of these organelles have not been investigated. Chlamydomonas pyrenoids are known to be liquid-like in nature, whereas the pyrenoids of the diatom *Phaeodactylum tricornutum* exhibit much less internal mixing [25]. To infer the physical properties of *A. agrestis* pyrenoids, we applied the fluorescence recovery after photobleaching (FRAP) technique on our stable transgenic line expressing Rubisco activase (RCA) tagged with mVenus. RCA was previously shown to co-localize with Rubisco in *A. agrestis* [23], and can thus be used as a pyrenoid marker. We found that the mVenus signal exhibited a rapid recovery ($t_{1/2}$ = 16.2 s) post photobleaching (Fig 1f,g). The unbleached region equilibrates with the bleached region of the same pyrenoid within 60 s (Fig. 1f,g), suggesting high levels of internal mixing. Furthermore, live cell imaging of dividing chloroplasts in the same RCA::mVenus line demonstrated elongation of a pyrenoid as it is being pulled to one of the daughter cells, again suggesting that *A. agrestis* pyrenoids are liquid-like (Fig. 1h, Movie S1).

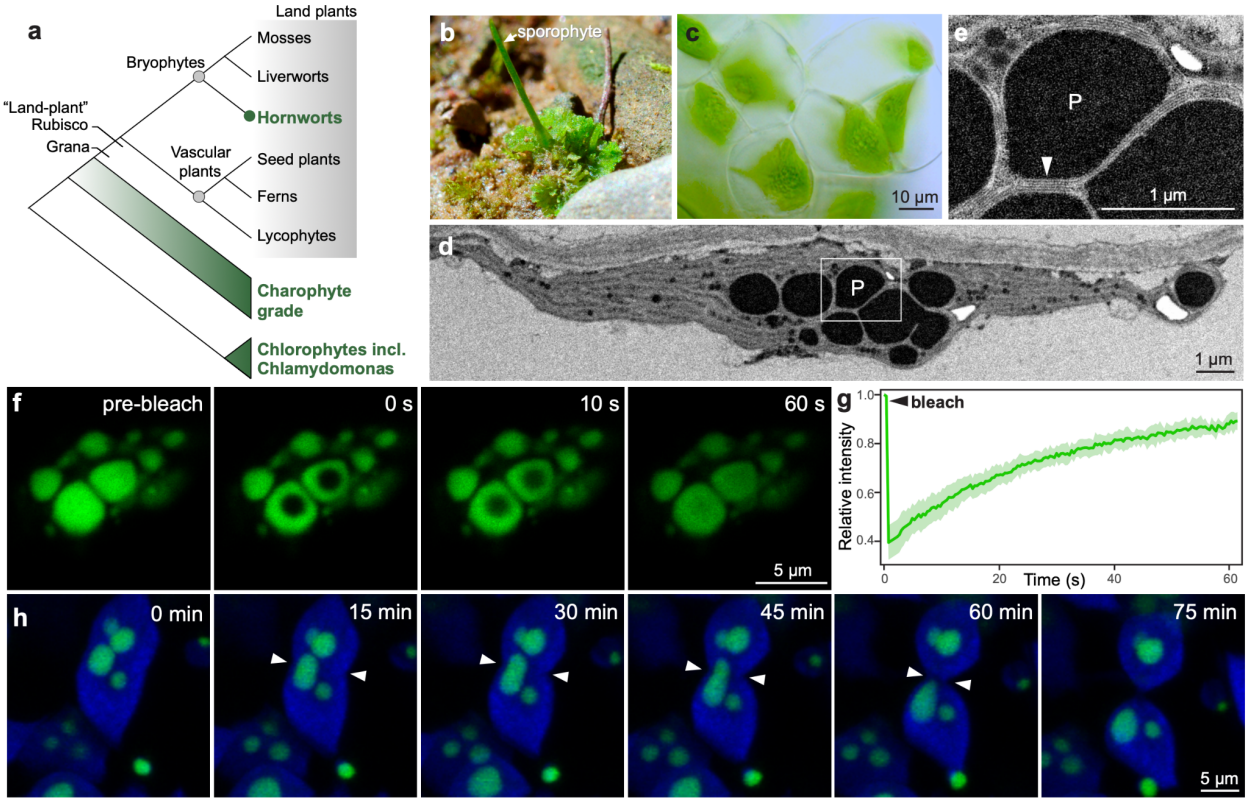

Figure 1. **Morphology and physical property of pyrenoids in *Anthoceros agrestis*.** (a) Hornworts are the only land plants that have a pCCM. Compared to Chlamydomonas, hornworts and crop plants share more common features in their chloroplast and Rubisco structures. (b) The model hornwort *Anthoceros agrestis* in native habitat. (c) *A. agrestis*, like many other hornwort species, has a single chloroplast per cell. (d) TEM image of an *A. agrestis* chloroplast with multiple pyrenoids, P. (e) A close-up view of a pyrenoid from (D). The arrowhead points to a stack of thylakoid membranes encasing the pyrenoid. (f) FRAP of hornwort pyrenoids labeled with RCA::mVenus. Photobleaching was targeted at the center of two selected pyrenoids. (scale bar, 5 µm.) (g) FRAP recovery curve. Error bars represent standard error of the mean (SEM) and are shown as the green shaded area (n=6). (h) Time-lapse images during cell division, showing signals of chlorophyll autofluorescence (blue) and RCA::mVenus fluorescence (green) over 75 mins. The cell division plane is demarcated with white arrowheads. (scale bar, 5 µm.)

Photosystems I and II are both in close proximity to pyrenoids

Given the tight association between thylakoids and pyrenoids in hornworts (Fig. 1d,e), we next examined how thylakoids are distributed across the chloroplasts of *A. agrestis*, especially whether there was differential localization of Photosystem I (PSI) and Photosystem II (PSII) relative to the pyrenoids. It is thought that in hornworts PSII localizes mainly to the grana, while PSI has a more ubiquitous distribution including in the channel thylakoids [20,26]. The evidence supporting this claim is, however, limited. To visualize PSI and PSII distribution, we transiently expressed fluorescently tagged Photosystem I reaction center subunit VI (PSAH) and Oxygen-evolving enhancer protein 2 (PSBP), respectively, in *A. agrestis* using a biolistics transformation approach [23].

In agreement with TEM images (Fig. 1d,e), confocal imaging of tagged PSAH and PSBP showed a tight association between thylakoids and pyrenoids (Fig. 2a,b; Supplementary Fig. 2). Pyrenoids are almost entirely enveloped by thylakoids (Fig. 2a,b), and both PSI and PSII localize in the membranes directly adjacent to the pyrenoids (Fig. 2a,b, d,e). Consistent with previous reports [20], PSI is uniformly distributed across the thylakoid network in hornworts, so much so that the signal is only absent in pyrenoid bodies (Fig. 2a; Supplementary Fig. 2). Localization of PSII appears to be less uniform than PSI, with gaps as well as smaller regions of much higher intensity, likely grana, scattered throughout the chloroplast (Fig. 2b; Supplementary Fig. 2).

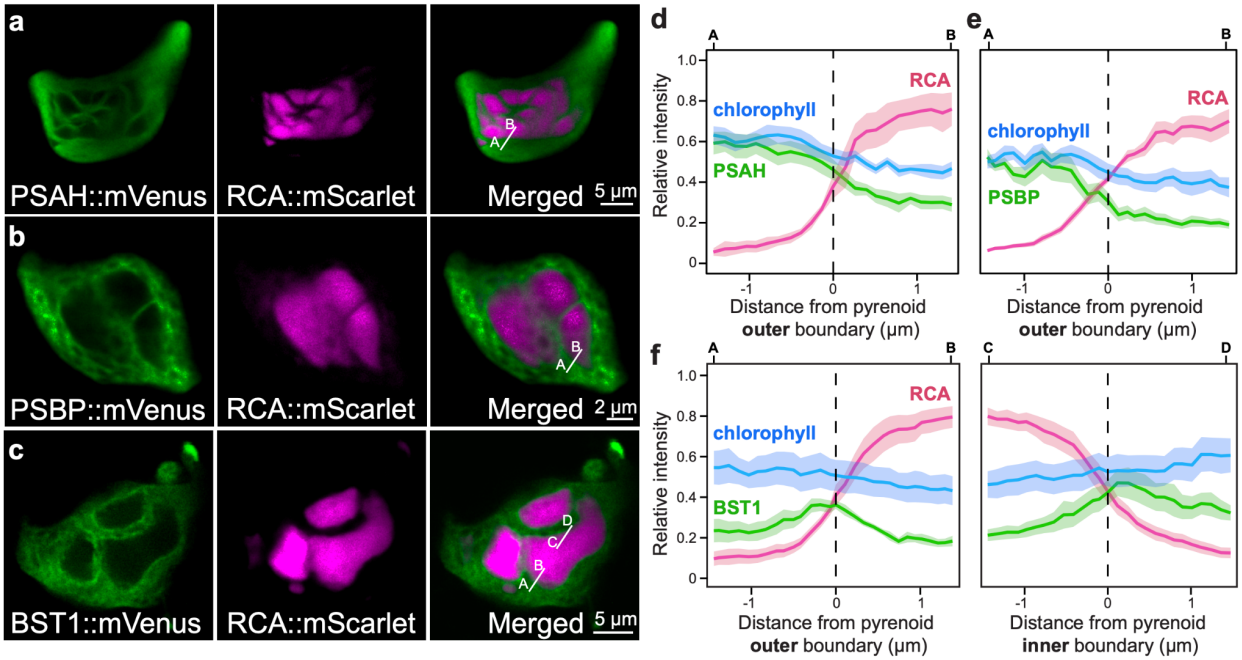

Figure 2. **Distribution of photosystems and BST channels on thylakoids of *A. agrestis*.** (a) Photosystem I (marked by PSAH::mVenus) and (b) photosystem II (marked by PSBP::mVenus) are distributed throughout the thylakoid network. (scale bar, 2 µm.) (c) BST1::mVenus has elevated fluorescent signals on the thylakoids immediately adjacent to pyrenoids. (scale bar, 5 µm). The fluorescent intensities of (d) PSAH::mVenus and (e) PSAH::mVenus decline along the transects going into pyrenoids (white lines in a and b). (f) Quantification of elevated BST1::mVenus fluorescent signals around pyrenoids (white lines in c). RCA::mScarlet was used as a pyrenoid marker. Error bars represent SEMs and are shown as shaded areas in each curve (n = 10, 11, and 7 for BST, PSII, and PSI, respectively).

Bestrophin channels localize to thylakoids adjacent to pyrenoids

A biophysical CCM in eukaryotes requires the presence of thylakoid membrane channels to allow HCO_3^- to enter the thylakoid lumen. In Chlamydomonas, a number of bestrophin channels (BST) fill this role [27]. There are four Chlamydomonas BST orthologs in the *A. agrestis* genome, but only one, BST1, is expressed abundantly (Supplementary Table 1). Fluorescent protein tagging indicated that BST1 in *A. agrestis* was distributed throughout the thylakoid network, but

exhibited elevated fluorescence intensity immediately adjacent to the pyrenoids (Fig. 2c,f; Supplementary Fig. 3). This pattern is in contrast to that of PSAH and PSBP, where the fluorescence intensities continuously decline going into pyrenoids (Fig. 2d,e). The localization of *A. agrestis* BST1 is similar to what was reported in Chlamydomonas [27], and implies that BST1 likely has the same function in hornworts, transporting HCO_3^- to thylakoid lumen.

β-CA1 localizes to the cytoplasm

We next examined the localization of several carbonic anhydrases (CA) in *A. agrestis*. There are three known CA families in land plants (α, β, and γ), which catalyze the reversible hydration of CO_2 to HCO_3^- [28]. While hornworts possess multiple CAs, we focused on three—β-CA1, CAH3, and LCIB—that have high potential to influence pCCM. β-CA1 is the second most abundant protein in plants after Rubisco and is typically located in the chloroplast stroma [29]. However, the presence of a CA in the stroma, and specifically the pyrenoid matrix, in hornwort chloroplasts would likely convert CO_2 into HCO_3^-, thus decreasing the availability of CO_2 for Rubisco and short circuiting the pCCM. The β-CA1 in *A. agrestis*, unlike the orthologs in C_3 plants, lacks a predicted chloroplast transit peptide. Fluorescently tagged *A. agrestis* β-CA1 confirmed the cytoplasmic localization (Supplementary Fig. 4). This absence of a diffuse stromal β-CA in *A. agrestis* is consistent with models of a pCCM [13] and reinforces that the relocalization of stromal CAs to cytoplasm is a key step toward engineering CCMs in crop plants [30].

CAH3 localizes to the space between pyrenoids

While a stroma-localized CA would prohibit pCCM function, treatment with a CA inhibitor (ethoxzolamide) greatly decreased the rate of photosynthesis and significantly increased the CO_2 compensation point of hornworts, but not liverworts [31], demonstrating that a CA is required for hornwort pCCM. In Chlamydomonas, CAH3 is located in thylakoid tubules, which form a knot in the center of the pyrenoid [32–34]. Chlamydomonas CAH3 converts lumenal HCO_3^- to CO_2, allowing for diffusion of CO_2 across the thylakoids and into the pyrenoid space. We discovered that the genomes of hornworts and some other bryophytes possess orthologs of the Chlamydomonas CAH3 gene, while vascular plants appear to have lost it (Fig. 3A; Supplementary Fig. 5). *A. agrestis* CAH3 is abundantly expressed, and retained not only the three critical histidine residues at the active site [35], but also the lumenal transit peptide (Supplementary Fig. 6, Supplementary Table 1). Importantly, we found that fluorescently tagged *A. agrestis* CAH3 formed puncta localized to the periphery of pyrenoids, concentrated adjacent (<500 nm) to the most interior side of the pyrenoids (Fig. 3b,c; Supplementary Fig. 6; see Movie S2 for Z-stack). CAH3 thus appears to be at the center of the chloroplast and in the space between pyrenoids. This localization pattern supports CAH3's role in supplying CO_2 to the surrounding pyrenoids in the hornwort pCCM.

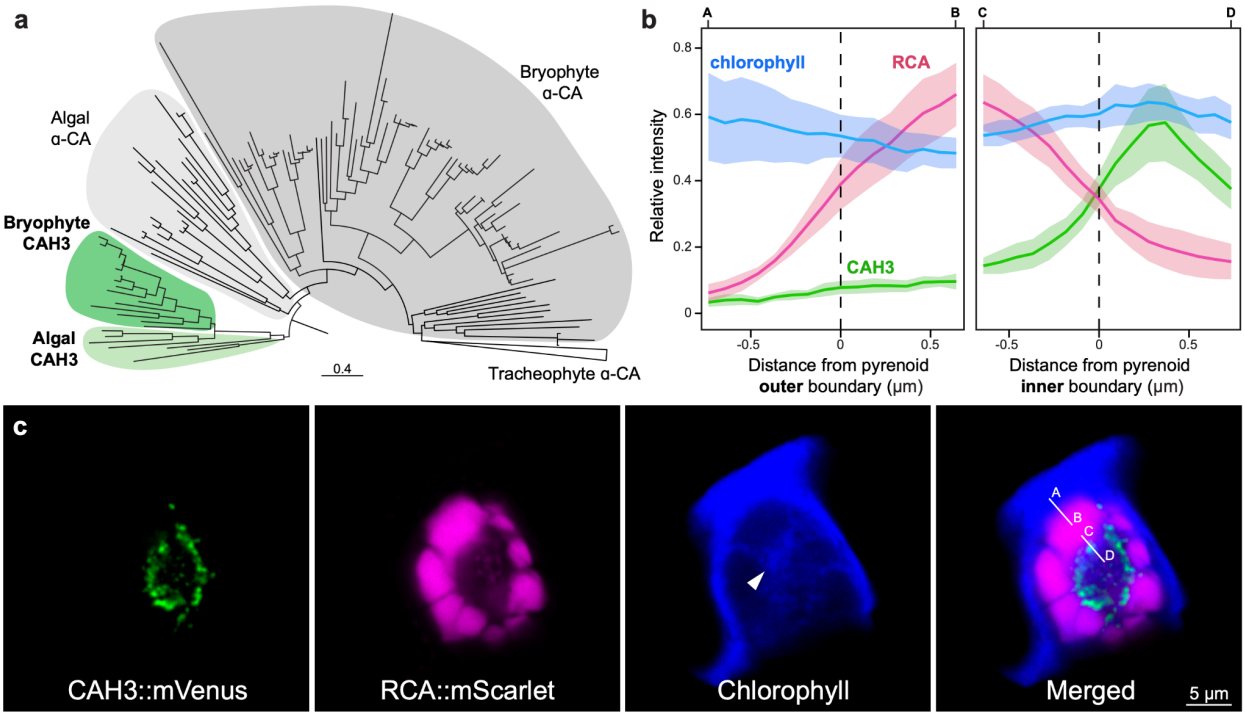

Figure 3. **Pyrenoids are organized around CAH3 in *Anthoceros agrestis*.** (a) Phylogeny of α-CA including CAH3. Orthologs of Chlamydomonas CAH3 can be found in hornworts and other bryophytes (shaded in green), but were lost in vascular plants. (b) CAH3::mVenus fluorescence intensity spikes at the interior side of pyrenoids (white lines in c). Error bars represent SEMs and are shown as shaded areas in each curve (n=10). (c) Example images of a cell co-transformed with CAH3::mVenus (green) and RCA::mScarlet (magenta). Chlorophyll autofluorescence is shown in blue. White arrowhead points to the central thylakoid knot. The Z-stack series can be found in Movie S2. (scale bar, 5 μm.)

LCIB localizes to the chloroplast membrane

Another CA crucial to Chlamydomonas pCCM is LCIB [36,37], which localizes to the periphery of Chlamydomonas pyrenoids under low CO_2 conditions to recapture leaked CO_2 from pyrenoids [38]. LCIB orthologs are present in hornworts with conserved residues for coordination of Zn^{2+} ion and catalysis [21]. Apart from hornworts, no LCIB homolog has been identified in other land plant genomes [21]. This exclusive presence of LCIB in pyrenoid-bearing algae and hornworts suggests that LCIB may have a role in hornwort pCCMs. We discovered that, unlike LCIB in Chlamydomonas, fluorescently tagged *A. agrestis* LCIB did not localize to the immediate pyrenoid periphery, but instead to the edge of the chloroplast (Supplementary Fig. 7). To interrogate this pattern further, we co-localized LCIB::mVenus with mScarlet tagged TIC40, which is a known component of the inner chloroplast membrane [39]. The fluorescence of LCIB::mVenus and TIC40::mScarlet clearly overlapped (Fig. 4), thus supporting that LCIB is indeed membrane-bound.

There are several reasons why the membrane-localized LCIB in *A. agrestis* is logical. In Chlamydomonas, LCIB localization to the starch sheath is likely mediated by its homolog LCIC [38]. Hornworts neither have starch sheaths nor LCIC, thus we would not expect to see a similar localization of LCIB. Furthermore, if LCIB were to surround each pyrenoid in hornworts, it would

likely rehydrate any CO_2 that had just been released by CAH3 (which are placed outside of pyrenoids; Fig. 3c), thus short-circuiting the pCCM. This scenario would be impossible in Chlamydomonas because CAH3 is localized in the thylakoid tubules embedded within a pyrenoid. Indeed, previous reaction-diffusion modeling of a pCCM showed that when thylakoid stacks are used as a diffusion barrier (as is the case of hornworts), LCIB localization to the chloroplast envelope would be optimal. This is because localizing LCIB as far from pyrenoids as possible minimizes "stealing" of CO_2 from the pyrenoid matrix and can still serve as the last line of defense for CO_2 leakage [13]. Furthermore, for any passively diffused CO_2 coming into the chloroplast, membrane-localized LCIB would quickly convert it into HCO_3^- before it can escape back to the cytoplasm.

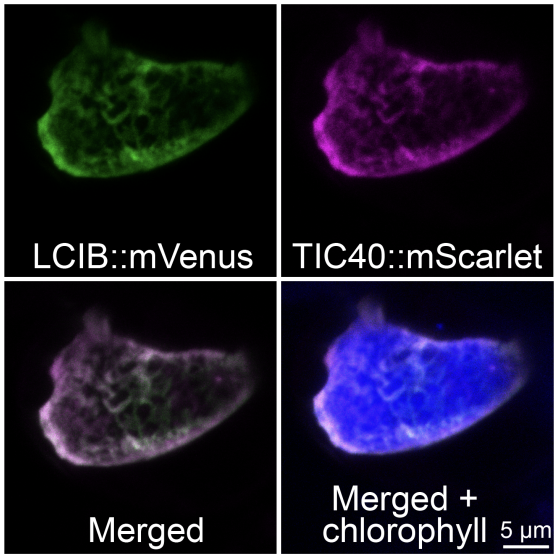

Figure 4. ***A. agrestis* LCIB localizes to the chloroplast membrane.** Maximum intensity projection of *A. agrestis* co-transformed with LCIB::mVenus (green) and TIC40::mScarlet (magenta). Chlorophyll autofluorescence is shown in blue. (scale bar, 5 µm.)

Co-immunoprecipitation and genome-scanning did not reveal putative linker proteins

Pyrenoids, which house Rubisco, are central to pCCM. Rubisco linker proteins that mediate pyrenoid formation have been identified in the green algae Chlamydomonas (EPYC1) [7] and *Chlorella sorokiniana* (CsLinker) [40], as well as in the diatom *Phaeodactylum tricornutum* (PYCO1) [25]. A putative linker was also proposed in the chlorarachniophyte alga *Amorphochlora amoebiformis* [41]. While there is little sequence similarity between these linker proteins, they share certain physicochemical properties, and importantly, can all be pulled down using co-immunoprecipitation (co-IP) with an anti-Rubisco antibody. To identify the possible Rubisco linker in *A. agrestis*, we optimized the lysis of *A. agrestis* tissue and conducted co-IP using a custom anti-Rubisco antibody with soluble proteins clarified from hornwort lysate (Supplementary Fig. 8). From our proteomic analysis we found that both Rubisco large and small subunits, as well as RCA were highly enriched in the immunoprecipitated fraction relative to the control where no antibodies were applied (Supplementary Fig. 9). Gene ontology terms relating to chloroplast, photosynthesis, and thylakoid membranes were significantly overrepresented in the immunoprecipitated sample, likely reflecting the tight association between pyrenoids and thylakoids (Supplementary Fig. 10). Together, these data suggest that

our co-IP approach was successful in enriching the bait (i.e. Rubisco) and its prey proteins (i.e. potential interactors).

Of the proteins significantly enriched by Rubisco co-IP, we searched for ones with similar physicochemical properties to other known Rubisco linker proteins. Namely, the candidate protein needs to: (1) be highly expressed, (2) be rich in repeat motif sequences, (3) be highly disordered, and (4) have a high isoelectric point (see Methods for detail). Only one protein was found (AnagrOXF.S1G403800.t1) possessing promising characteristics as a linker protein, which we named Putative Pyrenoid Protein 1 (PPP1). In addition to having an oscillating disorder profile, a chloroplast transit peptide, and high enrichment in the co-IP fraction, PPP1 has similarity to Structural Maintenance of Chromosomes (SMC) proteins (Supplementary Fig. 11). We found that fluorescently tagged PPP1 did not localize to the pyrenoid, but instead to the stroma (Supplementary Fig. 11). Interestingly, in some instances its localization became more concentrated in a region central to the chloroplast (Supplementary Fig. 11), reminiscent of the CAH3 localization pattern (Fig. 3). In any case, it is clear that PPP1 is not a Rubisco linker protein. Despite that our co-IP experiment found no evidence for a canonical linker in *A. agrestis*, we cannot rule out the possibility that one exists which has evaded our notice. Therefore, we broadened our search to include the entire predicted proteome of *A. agrestis*, using the same screening criteria as before. Only one candidate protein, PPP2 (AnagrOXF.S4G428300.t2), was found meeting these criteria. Compared to the known Rubisco linkers (EPYC1, CsLinker, and PYCO1), the repeats of PPP2 were much more irregular and spaced erratically across the protein sequence (Supplementary Fig. 12). Using fluorescent protein tagging, we showed that PPC2 did not localize to the pyrenoid, but instead to the spaces between pyrenoid subunits (Supplementary Fig. 12). Similar to PPP1, no evidence supports PPP2 being a linker protein for *A. agrestis* Rubisco. Our results thus point to a possibility that *A. agrestis* might employ a different mode of pyrenoid condensation from what has been reported from algae.

A. agrestis **pyrenoids contain proteins involved in Rubisco assembly and Calvin-Benson-Bassham cycle**

While the Rubisco linker in *A. agrestis* remains elusive, we found other proteins of interest that were enriched by immunoprecipitation. For example, Rubisco assembly factors (Raf1, Cpn60α, Cpn60β) showed significant enrichment by co-IP (Supplementary Fig. 9). We tested the localization of Cpn60β, a chaperonin required for Rubisco folding [42,43] by fluorescent protein-tagging, and found Cpn60β::mVenus to be clearly pyrenoid-localized (Fig. 5a, Supplementary Fig 13). To further explore if Rubisco biogenesis might occur within pyrenoids, we tagged another Rubisco assembly factor, RbcX2 [44]. While RbcX2 was not co-immunoprecipitated, it is also localized to the pyrenoids (Fig. 5b, Supplementary Fig 13).

Another protein enriched by co-IP was glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Chloroplastic GAPDH catalyzes the reductive step of the Calvin-Benson-Bassham (CBB) cycle and is composed of A and B subunits, both of which were significantly enriched in the immunoprecipitated fraction (Supplementary Fig. 9). To examine if GAPDH is a pyrenoid component, we fluorescently tagged the B subunit (GapB) and confirmed its pyrenoid localization (Fig. 5C, Supplementary Fig 14). It is possible that the CBB cycle (or part of it) takes

place in pyrenoids to limit the diffusion time of CBB cycle intermediates. Contrastingly, in Chlamydomonas, CBB enzymes are localized to the periphery of the pyrenoid [45].

Based on our results, it is plausible that *A. agrestis* pyrenoids are an inclusive compartment, where multiple cellular processes revolving around Rubisco take place. Future studies systematically tagging Rubisco assembly and CBB cycle proteins are needed to further test this hypothesis.

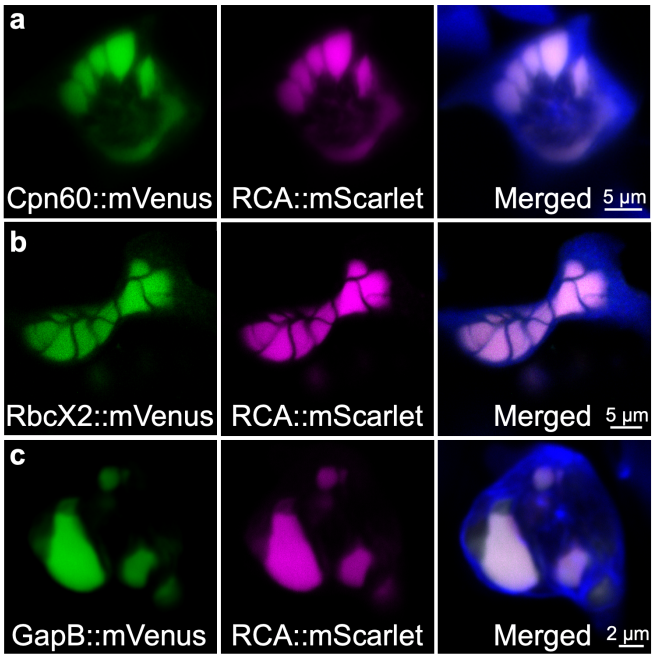

Figure 5. ***A. agrestis* Rubisco assembly and CCB cycle proteins localize to pyrenoids.** Example images of cells expressing either (a) Cpn60::mVenus, (b) RbcX2::mVenus, or (c) GapB::mVenus tagged with mVenus (green). RCA::mScarlet was co-transformed to mark pyrenoids (magenta). The merged images also included chlorophyll autofluorescence (blue). (scale bar, 2 or 5 µm.)

DISCUSSIONS

A working model for a hornwort pCCM

Our current evidence suggests a model (Fig. 6) for the hornwort pCCM that bears similarity to the algal pCCM yet with some key differences. First, since hornworts largely do not exist in an aqueous environment (where CO_2 diffusion is severely limited), we suspect that periplasmic CAs and active HCO_3^- transporters such as LCI1/HLA are not essential. Likewise, despite retaining the LCIB-BST-CAH3 chassis, hornworts do not have an ortholog to LCIA, the hypothesized HCO_3^- channel in Chlamydomonas. Given this, we hypothesize that the hornwort CCM likely operates under a “passive”, LCIB-dependent mode as described by Fei et al [13]. Once CO_2 passively diffuses across the chloroplast membrane it is rapidly converted to HCO_3^- by membrane-localized LCIB, trapping it in the stroma. In the stroma, HCO_3^- can enter the thylakoid space via BST1, where a concentration gradient drives it towards the pyrenoid adjacent thylakoids enriched in CAH3. There, CAH3 catalyzes the dehydration of HCO_3^-, creating a CO_2 enriched environment for the nearby pyrenoids. Leakage of CO_2 from the pyrenoid matrix is hindered by the highly reticulated thylakoid network, and any escaped CO_2 could be recaptured by LCIB at the chloroplast membrane. While we cannot rule out the presence of active Ci transport, the condition to drive an efficient pCCM is met based on the

localizations of LCIB, BST, and CAH3, thylakoid-enclosed pyrenoids, as well as the reaction-diffusion model by Fei et al [13].

The spatial model we presented here provides a framework for future hypothesis-driven research. Although protein localization alone does not provide the full proof for a protein's function, currently no reverse genetics method has been successfully developed in *A. agrestis*. Future work investigating carbon assimilation efficiency of selective gene knockdown or knockout mutants, as well as biochemical *in vitro* characterizations will be crucial to confirm or revise our model.

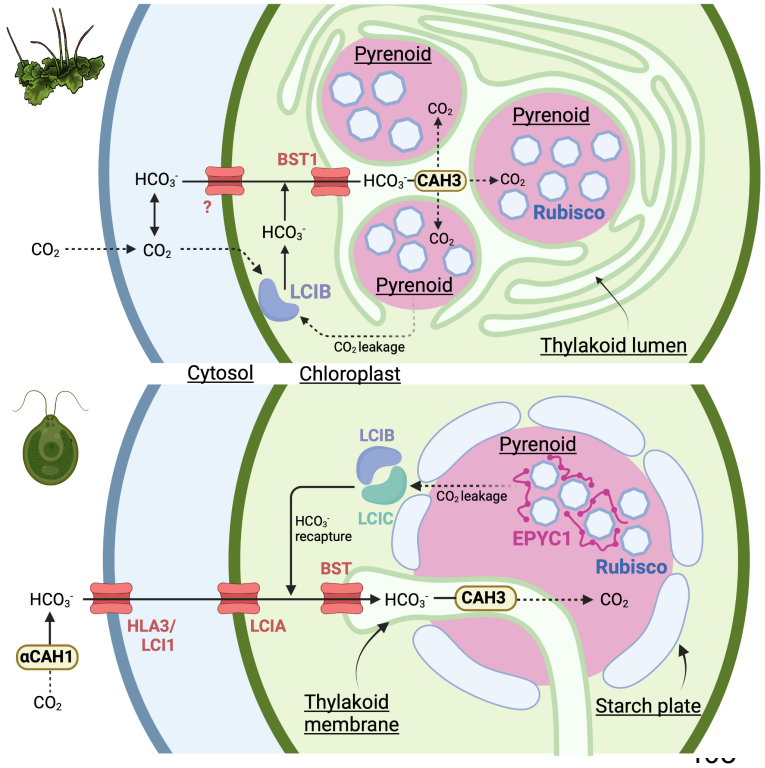

Figure 6. **The spatial model of *A. agrestis* pCCM and (top) comparison with that of Chlamydomonas (bottom).** The hornwort illustration was done by Rebecca Key.

The first land plants likely had the chassis for a biophysical CCM

Operating an effective biophysical CCM requires the strategic placement of Ci channels and CAs. It is striking that some of these key components (LCIB, BST, and CAH3) characterized in Chlamydomonas have orthologs in hornworts and likely play similar roles in their respective pCCM. It is difficult, if not impossible, to determine whether hornwort pCCMs represent an ancestral trait or an independent origin. In either case, however, the presence of the LCIB-BST-CAH3 chassis in hornworts suggests that the underlying chassis to concentrate CO_2 was present in the last common ancestor of land plants. Considering that various algae possess biophysical CCMs even in the absence of pyrenoids [46–49], it is not impossible that the earliest land plants could have operated a functional CCM.

We propose two scenarios that could result in hornworts being the only land plants having a pCCM. In one, pyrenoids were *de novo* evolved in hornworts on top of the LCIB-BST-CAH3 chassis already in place. The loss of LCIB and CAH3 in other plant lineages rendered the evolution of pyrenoids less advantageous. The supposed absence of a canonical Rubisco linker

for pyrenoid formation in *A. agrestis* lends support for the separate origins of pyrenoids. Alternatively, pyrenoids could be ancestral to land plants, but were subsequently lost once in setaphytes (mosses and liverworts) and another time in vascular plants. In this scenario hornworts did not independently evolve pyrenoids, but rather retained them as a plesiomorphic trait.

Implications for the repeated evolution of pyrenoids within hornworts

Within hornworts, pyrenoids have been lost and gained multiple times across the phylogeny [50]. Interestingly, LCIB, BST, and CAH3are present in all the sequenced hornwort species, even in those lacking pyrenoids [51]. Considering that pyrenoid-absent species have CO_2 compensation points which are intermediate between liverworts and pyrenoid-containing hornworts [52], a baseline biophysical CCM probably exists among most hornworts and the strength of this CCM could be accentuated by the presence of a pyrenoid. The universal presence of this LCIB-BST-CAH3 chassis could explain the repeated gains of pyrenoids as it might have predisposed the evolution of pyrenoids. The specific localization of CAs and Ci channels is also likely the first step to the evolution of a fully functional pCCM, rather than pyrenoid formation itself.

Hornworts provide a blueprint for engineering a biophysical CCM in land plants

Significant strides have been made in transplanting parts of the algal CCM to Arabidopsis in recent years. For example, by introducing EPYC1, SAGA1, and SAGA2 proto-pyrenoids with starch sheaths have been successfully built in Arabidopsis [53,54]. However, the entire algal CCM module has yet to be integrated into flowering plants. Our investigation into the hornwort system demonstrates that some components of the algal pCCM may not be required in land plants. First, a starch sheath around pyrenoids may not be necessary as thylakoid stacks could provide a sufficient CO_2 diffusion barrier (Fig. 1). Second, if hornworts indeed operate a “passive” pCCM, then our work provides the first empirical evidence for its feasibility [13]. In a passive pCCM, the proper localization of a minimal set of carbonic anhydrases (e.g. LCIB and CAH3) and thylakoid localized Ci channels (e.g. BST) may be sufficient to concentrate CO_2 without the need for active Ci transport to the chloroplast [13]. Importantly, recent modeling suggested that introducing active HCO_3^- transporters to build pCCMs in land plants is only viable if chloroplast membrane permeability to CO_2 is low [55]. The passive pCCM of hornworts would not encounter this challenge and could be potentially simpler to engineer. Taken together, hornworts provide an alternative blueprint to design a future pCCM.

METHODS

Anthoceros agrestis cultures

Anthoceros agrestis axenic cultures were maintained on *A. agrestis* gametophyte growth medium (AG medium [56]) supplemented with 0.2% sucrose following the protocol described in Lafferty et al [23]. The growth conditions included a temperature of 22°C, a 16/8 h light/dark cycle, and light intensity ranging from 6 to 25 µmol/m²/s. Warm white and cool white light was emitted by Ecolux XL Starcoat F32T8 XL SP30 ECO fluorescent bulbs and F32T8 XL SP41 ECO bulbs, respectively (General Electric, USA). A total of 3-6 g of fresh thallus tissue was homogenized and prepared for transformation as described in Lafferty et al [23]. The homogenate was filtered using a 100 µM cell strainer (MTC Bio, USA), and the thallus tissue was washed with sterile deionized water and plated on AG medium supplemented with 2% sucrose.

Transmission electron microscopy

Plant materials were processed for transmission electron microscopy following the published protocol [57], with some modifications. Pure cultures were fixed in 3% glutaraldehyde, 1% formaldehyde, and 0.75% tannic acid in 0.05 M Na-cacodylate buffer, pH 7, for 3-4 h at room temperature. After several rinses in Sorensen's 0.1 M buffer, the samples were post-fixed in 2% osmium tetroxide, dehydrated in an ethanol series and embedded in Spurr's resin for 24 h. Thin sections were cut with a diamond knife, stained with methanolic uranyl acetate for 15 min followed by Reynolds' lead citrate for 10 min, and observed with a Hitachi H-7100 transmission electron microscope at the Imaging-Microscopy Platform of the IBIS, Université Laval.

Construction of fluorescently tagged protein expression cassettes

Constructs were designed and built using the OpenPlant toolkit [58]. All genes of interest were driven by the same constitutive promoter (*A. agrestis* native *Elongation factor 1 alpha* (*Ef1α*) promoter) and *Nos* terminator (OP-053). Coding sequences of target genes were synthesized by Twist Biosciences with synonymous substitutions made where necessary to remove internal *Bsa*I or *Sap*I Type II restriction cut sites or to reduce repetitive nucleotide sequences if required for synthesis. Assembly reactions were performed by a one-pot mixture that consists of Type II restriction enzymes with T4 DNA ligase (NEB, USA) as described in Sauret-Gueto et al. [58]. Heat shock transformation with NEB® 5-alpha Competent *Escherichia coli* was used to regenerate ligated products on the respective Luria Broth agar plates with either kanamycin or streptomycin. Colony PCR was performed with KOD One™ PCR Master Mix (Toyobo) to screen colonies before recovering plasmids with a 5 mL overnight culture and miniprep (QIAgen). Whole plasmid sequencing (Eurofin Genomics, USA) was performed as a final validation. To prepare plasmids for biolistic mediated transformation, a total of 50 ml *E. coli* culture was grown overnight at 37 °C. The plasmid DNA was extracted using the PureYield™ Plasmid Midiprep system (Promega, USA), followed by vacuum concentration in a SpeedVac Concentrator (Thermofisher Scientific, USA) to obtain DNA concentrations of 1 µg/µL.

Transient expression of fluorescently tagged proteins in *A. agrestis*

To obtain transient expression of fluorescently tagged proteins in *A. agrestis*, we performed biolistic mediated transformation using DNA/gold particle mixture preparations and the particle Delivery system PDS-1000/He (Bio-Rad, USA) as described in Lafferty et al [23], with a target

distance of 14 cm and a burst pressure of 450 psi, under a vacuum of 28 inHg. The bombarded tissue was left to recover for 3-10 days, under standard culturing conditions as described above, before imaging.

Confocal imaging

Imaging was performed using Zeiss LSM710, Leica TCS SP5 Laser Scanning Confocal Microscope, or Leica Stellaris 5 confocal microscope. Proteins tagged with mVenus were excited with a 514 nm laser line with an emission band of 524/551 nm. Chlorophyll autofluorescence and proteins tagged with mScarlet were excited with a 561 nm laser line, with emission bands of 658/699 and 579/609 nm, respectively. Each FRAP experiment started with four initial scans before bleaching the region of interest (ROI). Circular ROIs with a 1.3 µm diameter were exposed eight times at 20% intensity of the 514 nm laser, before allowing recovery for 160 cycles (66 seconds) with laser power attenuated at 0.5% intensity. Replicates were collected by photobleaching individual pyrenoids within a cell. Subsequent analysis was conducted using the FRAP wizard in Leica Application Suite Advanced Fluorescence and data was plotted in R. Intensity profiles for all tagged proteins were collected by using the line tool to transect regions of interest (e.g. pyrenoid boundaries, chloroplast boundaries) and then plotting the intensity profile in Fiji [59]. These intensity profiles were then exported for subsequent analyses and plot generation in R.

Homology-based approach to identify candidate CCM proteins

To identify candidate CCM genes in hornworts we used Orthofinder v2.5.4 on a broad sampling of plant and algal genomes (Supplemental Table 2), including 11 species of hornworts [51], to generate orthogroups. We then selected orthogroups containing key Chlamydomonas CCM genes (Supplemental Table 1). Because orthogroups can sometimes represent large gene families, to infer the direct orthologues in hornworts, phylogenetic reconstruction using IQ-TREE [60] was performed to identify the hornwort sequences in the same subfamily as the Chlamydomonas genes of interest. From these, we selected *A. agrestis* genes which showed high expression levels using the RNA-seq data from Li et al [21]. In cases where there were multiple orthologues present in *A. agrestis*, we selected the gene with the highest expression. In addition to CCM genes, we choose Photosystem I reaction center subunit VI (PSAH) and Oxygen-evolving enhancer protein 2 (PSBP) as Photosystem I and Photosystem II markers, respectively, to compare the distribution of photosystems and to visualize thylakoids.

Targeted bioinformatic search for a Rubisco linker protein

A targeted search for potential linker proteins was conducted in a manner similar to, but less stringent than described in Mackinder et al [7]. Genomic sequences of two *A. agrestis* strains [21] were first screened for tandem repeats using Xstream [61]. Default parameters were used with the following exceptions: Minimum Period: 20; Maximum Period: 100; Minimum Copy # 2; and Minimum tandem repeat Domain: 40. Proteins which passed this screen were then analyzed for: 1) a high isoelectric point (>8) using Expasy [62], 2) a high proportion of disordered sequence using the PONDR server (>50%) [63]. From this screen, lowly expressed genes were filtered based upon their corresponding RNA transcript abundance (transcript per million <10) [21].

Optimizing lysis of *A. agrestis* thallus

To derive soluble proteins from *A. agrestis* thallus, we experimented the lysis of *A. agrestis* by testing four different buffers: buffer A (20 mM HEPES pH 7.0, 50 mM NaCl, 12.5% v/v glycerol), buffer B (Buffer A with 2% v/v Triton X-100), buffer C (20 mM CAPS pH 11.0, 50 mM NaCl, 12.5% v/v glycerol), and buffer D (Buffer C supplemented with 2% v/v Triton X-100). For each lysis, 5 g of *A. agrestis* tissue were harvested from liquid cultures in the AG medium [56]. The biomass was frozen with liquid N_2 and crushed to find powder with a mortar and pestle. A total of 5 mL of either test buffer supplemented with 1 mM phenylmethylsulfonyl fluoride and one protease inhibitor tablet (Roche) was then to resuspend the fine powder. The suspension was lysed with a French pressure cell press (American Instrument Company, USA) at 1000 PSI. Polyvinylpolypyrrolidone (PVPP, 2%) was added to the French pressed lysate and the mixture was sieved through four layers of Miracloth, pre-wetted with the respective test buffer. The filtrate was centrifuged (21,000 *g* for 45 min at 4 °C) to separate pellet and supernatant fractions. The supernatant fraction was further filtered through a 0.22 µm syringe filter. Both fractions were analyzed on SDS-PAGE and western blot using 8–16% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad). For immunoblotting of Rubisco, a polyclonal antibody was raised in rabbits against the *A. agrestis* Rubisco large subunit C-terminus peptide EVWKEIKFVFETIDTL and affinity purified (Life Technologies, USA).

Co-immunoprecipitation of Rubisco

A total of 100 µL Protein A resin (Dynabeads, Invitrogen) was aliquoted into 1.5 mL microcentrifuge tubes and washed thrice with buffer C (20 mM CAPS pH 11.0, 50 mM NaCl, 12.5% v/v glycerol) with a magnetic rack. Next, the resin was incubated with 7.5 µg/mL anti-Rubisco on a rotator (2 h/4°C) and washed twice with buffer C. Controls were performed by not priming the resin with anti-Rubisco antibodies. A total of 750 µL soluble lysate was then added to anti-Rubisco bound resin and incubated on a rotator overnight (~18 h/4 °C) and washed thrice with buffer C. Protein elution was carried out by adding 80 µL of 2.5X SDS loading buffer, without 2-Mercaptoethanol. Eluted proteins were separated from the Dynabeads and 2-Mercaptoethanol was added to a final concentration of 100 mM, proteins were then boiled (95 °C/5 min) and shipped for LC-MS/MS analysis. The co-IP and control experiments each had five technical replicates.

Proteomic analysis

LC-MS/MS analysis was performed at the Environmental Molecular Sciences Laboratory in Pacific Northwest National Laboratory. Samples were processed using Filter Aided Sample Preparation (FASP) [64] by adding 400 µl of 8 M urea to 30K molecular weight cut off (MWCO) FASP spin columns along with 40 µl of the sample in SDS BME buffer and centrifuged at 14,000 *g* for 20 min. Urea washes were repeated three additional times followed by the addition 400 µl of 50 mM ammonium bicarbonate, pH 8.0 and two repeated centrifugation for 20 min. The columns were then placed into clean and labeled collection tubes. The digestion solution was made by dissolving 5 µg trypsin in 75 µL 50 mM ammonium bicarbonate solution which was added to each sample. The samples were then incubated for 3 h at 37 °C with 600 rpm shaking on a thermomixer with a ThermoTop (Eppendorf, Hamburg, Germany) to reduce condensation

into the caps of collection tubes. The resultant peptides were then centrifuged through the filter and into the collection tube by centrifuging at 14,000 *g* for 15 mins. The peptides were concentrated to ~30 µL using a vacuum concentrator. Final peptide concentrations were determined using a bicinchoninic acid (BCA) assay (Thermo Scientific, Waltham, MA USA) and each sample was prepared at 0.1 µg/µl for MS analysis

Digested protein samples were analyzed using an Orbitrap Eclipse Tribrid MS (Thermo Scientific) outfitted with a high field asymmetric waveform ion mobility spectrometry (FAIMS) interface, using data-dependent acquisition mode. Peptides were ionized using a voltage of 2.4 kV and with an ion transfer tube temperature at 300 °C. Data acquisition time was 2 hr following a 20 min delay to avoid dead time between injection and elution of peptides. A proprietary method for transferring identification based on FAIMS filtering was used to fractionate ionized peptides by the FAIMSpro interface using a 3-Compensation Voltage (3-CV); -45, -60, -75 V method. Fractionated ions with a mass range 400-1800 m/z were scanned with Orbitrap at 120,000 resolution with an injection time (IT) of 50 ms and an automatic gain control (AGC) target of 4e5. Cycle times of 1.0 s were used for the 3-CV method. Precursor ions with intensities > 1e4 were fragmented with an isolation window of 0.7 by 30% higher-energy collisional dissociation energy and scanned with an AGC target of $1e^4$ as well as an IT of 35 ms.

Raw data files were referenced to *A. agrestis* nuclear encoded and chloroplast encoded proteins [21], and peptide abundances were extracted from the raw spectra using MASIC [65] and log2 transformed to remove skewness in distribution of measured abundances. Transformed abundance values were then normalized using the mean central tendency method implemented in InfernoR [66]. Normalized peptide abundances were de-logged, summed, transformed (log2), and normalized again in InfernoR to produce normalized abundances for the protein level roll-up. Proteins which were missing in more than one replicate of the conditions (control or anti-RbcL) were filtered from the final analysis to limit the imputation of too many missing values. Left-censored missing values were imputed using the Minimum Probability method with the default parameters. Differential enrichment analysis was conducted using the DEP Bioconductor package version 1.27.0 [67] with a p value cutoff of 0.05. Plots were generated using ggplot2 (version 3.5.1). Gene ontology (GO) enrichment analysis was carried out using the GO Enrichment module of TBtools [68] with *goslim_plant* selected. Background file was set as the entire *A. agrestis* proteome and proteins found to be significantly enriched by coimmunoprecipitation were chosen as the selection set. The resulting table was then used to generate an enrichment barplot.

ACKNOWLEDGEMENTS

This work is supported by National Science Foundation MCB-2213841 to F.-W.L. and MCB-2213840 to L.H.G, Environmental Molecular Sciences Laboratory User Grant to F.-W.L., Triad Foundation Grant to F.-W.L., and Schmittau-Novak Graduate Student Grant to T.A.R. We thank Aleksandra Skirycz, Kathryn Eshenour, Amber Hotto, and David Stern for providing access to tools and reagents to establish preliminary lysis and co-IP experiments, Rebecca Key for the hornwort illustration, Carrie Nicora and Nikola Tolic from Environmental Molecular Sciences Laboratory in Pacific Northwest National Laboratory for technical assistance with mass spectrometry processing, the Plant Cell Imaging Center at Boyce Thompson Institute and Mamta Srivastava for technical assistance to confocal imaging with Leica TCS SP5 Laser Scanning Confocal Microscope, Adrienne Roeder for providing access to Zeiss LSM710 and Leica Stellaris 5 confocal microscope, the York Physics of Pyrenoids research community for feedback, and Li and Gunn lab members for discussions.

AUTHOR CONTRIBUTIONS

T.A.R., L.H.G., and F.-W.L conceived the project. T.A.R. made the gene constructs and carried out confocal imaging. T.A.R., D.L., and X.X. performed hornwort transformation. Z.G.O. designed the Rubisco antibody and optimized the lysis of hornwort thallus. T.A.R. and Z.G.O. performed co-immunoprecipitation. J.C.A.V. performed transmission electron microscopy. T.A.R., Z.G.O., L.H.G. and F.-W.L. analyzed the data. T.A.R. and F.-W.L. wrote the manuscript with contributions and comments from all authors. L.H.G. and F.-W.L. secured the funding and supervised the project.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Bar-On, Y. M. & Milo, R. The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4738–4743 (2019).
2. Ellis, R. J. The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244 (1979).
3. Leegood, R. C. A welcome diversion from photorespiration. *Nat. Biotechnol.* **25**, 539–540 (2007).
4. Long, S. P., Zhu, X.-G., Naidu, S. L. & Ort, D. R. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* **29**, 315–330 (2006).
5. Griffiths, H., Meyer, M. T. & Rickaby, R. E. M. Overcoming adversity through diversity: aquatic carbon concentrating mechanisms. *J. Exp. Bot.* **68**, 3689–3695 (2017).
6. Badger, M. R. & Price, G. D. CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622 (2003).
7. Mackinder, L. C. M. *et al.* A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5958–5963 (2016).
8. Freeman Rosenzweig, E. S. *et al.* The eukaryotic CO2-concentrating organelle is liquid-like and exhibits dynamic reorganization. *Cell* **171**, 148–162.e19 (2017).
9. Wunder, T., Cheng, S. L. H., Lai, S.-K., Li, H.-Y. & Mueller-Cajar, O. The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat. Commun.* **9**, 5076 (2018).

10. He, S., Crans, V. L. & Jonikas, M. C. The pyrenoid: the eukaryotic CO2-concentrating organelle. *Plant Cell* **35**, 3236–3259 (2023).
11. Ramazanov, Z. *et al.* The induction of the CO2-concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of Chlamydomonas reinhardtii. *Planta* **195**, (1994).
12. Itakura, A. K. *et al.* A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18445–18454 (2019).
13. Fei, C., Wilson, A. T., Mangan, N. M., Wingreen, N. S. & Jonikas, M. C. Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nature Plants* **8**, 583–595 (2022).
14. McGrath, J. M. & Long, S. P. Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol.* **164**, 2247–2261 (2014).
15. Yin, X. & Struik, P. C. Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *J. Exp. Bot.* **68**, 2345–2360 (2017).
16. Irisarri, I. *et al.* Unexpected cryptic species among streptophyte algae most distant to land plants. *Proceedings of the Royal Society B* **288**, 20212168 (2021) .
17. Vaughn, K. C., Campbell, E. O., Hasegawa, J., Owen, H. A. & Renzaglia, K. S. The pyrenoid is the site of ribulose 1, 5-bisphosphate carboxylase/oxygenase accumulation in the hornwort (Bryophyta: Anthocerotae) chloroplast. *Protoplasma* **156**, 117–129 (1990).
18. Smith, E. & Griffiths, H. A pyrenoid-based carbon-concentrating mechanism is present in terrestrial bryophytes of the class Anthocerotae. *Planta* **200**, (1996).
19. Li, F.-W., Villarreal Aguilar, J. C. & Szövényi, P. Hornworts: An overlooked window into carbon-concentrating mechanisms. *Trends Plant Sci.* **22**, 275–277 (2017).
20. Vaughn, K. C. *et al.* The anthocerote chloroplast: a review. *New Phytol.* **120**, 169–190 (1992).
21. Li, F.-W. *et al.* Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat Plants* **6**, 259–272 (2020).
22. Szövényi, P. *et al.* Establishment of Anthoceros agrestis as a model species for studying the biology of hornworts. *BMC Plant Biol.* **15**, 98 (2015).
23. Lafferty, D. J. *et al.* Biolistics-mediated transformation of hornworts and its application to study pyrenoid protein localization. *J. Exp. Bot.* erae243 (2024).
24. Barrett, J., Girr, P. & Mackinder, L. C. M. Pyrenoids: CO2-fixing phase separated liquid organelles. *Biochim. Biophys. Acta Mol. Cell Res.* **1868**, 118949 (2021).
25. Oh, Z. G. *et al.* A linker protein from a red-type pyrenoid phase separates with Rubisco via oligomerizing sticker motifs. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2304833120 (2023).
26. Burr, F. A. Phylogenetic Transitions in the Chloroplasts of the Anthocerotales. I. The Number and Ultrastructure of the Mature Plastids. *Am. J. Bot.* **57**, 97–110 (1970).
27. Mukherjee, A. *et al.* Thylakoid localized bestrophin-like proteins are essential for the CO2 concentrating mechanism of Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences* **116**, 16915–16920 (2019).
28. Ignatova, L., Rudenko, N., Zhurikova, E., Borisova-Mubarakshina, M. & Ivanov, B. Carbonic Anhydrases in Photosynthesizing Cells of C3 Higher Plants. *Metabolites* **9**, 73 (2019).
29. Badger, M. R. & Price, G. D. The role of carbonic anhydrase in photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 369–392 (1994).
30. Jethva, J. *et al.* Realisation of a key step in the evolution of C4 photosynthesis in rice by genome editing. *bioRxiv* 2024.05.21.595093 (2024).
31. Smith, E. C. & Griffiths, H. The role of carbonic anhydrase in photosynthesis and the activity of the carbon-concentrating-mechanism in bryophytes of the class Anthocerotae.

New Phytol. **145**, 29–37 (2000).
32. Markelova, A. G., Sinetova, M. P., Kupriyanova, E. V. & Pronina, N. A. Distribution and functional role of carbonic anhydrase Cah3 associated with thylakoid membranes in the chloroplast and pyrenoid of Chlamydomonas reinhardtii. *Russ. J. Plant Physiol.* **56**, 761 (2009).
33. Sinetova, M. A., Kupriyanova, E. V., Markelova, A. G., Allakhverdiev, S. I. & Pronina, N. A. Identification and functional role of the carbonic anhydrase Cah3 in thylakoid membranes of pyrenoid of Chlamydomonas reinhardtii. *Biochim. Biophys. Acta* **1817**, 1248–1255 (2012).
34. Mitra, M. *et al.* The carbonic anhydrase gene families of Chlamydomonas reinhardtii. *Can. J. Bot.* **83**, 780–795 (2005).
35. Terentyev, V. V. & Shukshina, A. K. CAH3 from Chlamydomonas reinhardtii: Unique Carbonic Anhydrase of the Thylakoid Lumen. *Cells* **13**, 109 (2024).
36. Wang, Y. & Spalding, M. H. LCIB in the Chlamydomonas CO2-concentrating mechanism. *Photosynth. Res.* **121**, 185–192 (2014).
37. Jin, S. *et al.* Structural insights into the LCIB protein family reveals a new group of β-carbonic anhydrases. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14716–14721 (2016).
38. Yamano, T., Toyokawa, C., Shimamura, D., Matsuoka, T. & Fukuzawa, H. CO2-dependent migration and relocation of LCIB, a pyrenoid-peripheral protein in Chlamydomonas reinhadtii. *Plant Physiol.* **188**, 1081–1094 (2021).
39. Chou, M.-L. *et al.* Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. *EMBO J.* **22**, 2970–2980 (2003).
40. Barrett, J. *et al.* A promiscuous mechanism to phase separate eukaryotic carbon fixation in the green lineage. *bioRxiv* 2024.04.09.588658 (2024).
41. Moromizato, R. *et al.* Pyrenoid proteomics reveals independent evolution of the CO_2-concentrating organelle in chlorarachniophytes. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2318542121 (2024).
42. Aigner, H. *et al.* Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).
43. Feiz, L. *et al.* Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012).
44. Saschenbrecker, S. *et al.* Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
45. Wang, L. *et al.* A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. *Cell* **186**, 3499–3518.e14 (2023).
46. Badger, M. R. *et al.* The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. *Can. J. Bot.* **76**, 1052–1071 (1998).
47. Kevekordes, K. *et al.* Inorganic carbon acquisition by eight species of Caulerpa (Caulerpaceae, Chlorophyta). *Phycologia* **45**, 442–449 (2006).
48. Gee, C. W. & Niyogi, K. K. The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism in Nannochloropsis oceanica. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4537–4542 (2017).
49. Steensma, A. K., Shachar-Hill, Y. & Walker, B. J. Correction to: The carbon-concentrating mechanism of the extremophilic red microalga Cyanidioschyzon merolae. *Photosynth. Res.* **158**, 203 (2023).
50. Villarreal, J. C. & Renner, S. S. Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18873–18878 (2012).
51. Schafran, P. *et al.* Pan-phylum genomes of hornworts revealed conserved autosomes but dynamic accessory and sex chromosomes. *Nat Plants* in review (2024).
52. Hanson, D., John Andrews, T. & Badger, M. R. Variability of the pyrenoid-based CO2

concentrating mechanism in hornworts (Anthocerotophyta). *Funct. Plant Biol.* **29**, 407–416 (2002).
53. Atkinson, N., Mao, Y., Chan, K. X. & McCormick, A. J. Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. *Nat. Commun.* **11**, 6303 (2020).
54. Atkinson, N., Stringer, R., Mitchell, S. R., Seung, D. & McCormick, A. J. SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in Arabidopsis chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2311013121 (2024).
55. Kaste, J. A. M., Walker, B. J. & Shachar-Hill, Y. Reaction-diffusion modeling provides insights into biophysical carbon concentrating mechanisms in land plants. *Plant Physiol.* kiae324 (2024).
56. Gunadi, A., Li, F.-W. & Van Eck, J. Accelerating gametophytic growth in the model hornwort Anthoceros agrestis. *Appl. Plant Sci.* **10**, e11460 (2022).
57. Villarreal, J. C., Duckett, J. G. & Pressel, S. Morphology, ultrastructure and phylogenetic affinities of the single-island endemic Anthoceros cristatus Steph. (Ascension Island). *J. Bryol.* **39**, 226–234 (2017).
58. Sauret-Gueto, S. *et al.* Systematic tools for reprogramming plant gene expression in a simple model, Marchantia polymorpha. *ACS Synth. Biol.* **9**, 864–882 (2020).
59. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
60. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
61. Newman, A. M. & Cooper, J. B. XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics* **8**, 382 (2007).
62. Gasteiger, E. *et al.* ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
63. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).
64. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
65. Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput. Biol. Chem.* **32**, 215–217 (2008).
66. Polpitiya, A. D. *et al.* DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **24**, 1556–1558 (2008).
67. Zhang, X. *et al.* Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.* **13**, 530–550 (2018).
68. Chen, C. *et al.* TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).